Mullins Consulting, Inc.
The Web Site of Craig S. Mullins

A Guide to Db2 Application Performance for Developers
By Craig S. Mullins
Order now!
A new book to help programmers write efficient Db2 code
Covers both Db2 for z/OS and LUW

Home   Services   Articles   Presentations   Books   Speaking 2021   Social Media   Database Links   Contact Us

Enterprise

Leadership

Gaining Knowledge Through Process Mining

by Craig S. Mullins

June 2006

A large wireless phone service provider was concerned with the number of customers it was losing. Every customer lost costs the company $53 in monthly revenue. Although the revenue looks small on a customer by customer basis, with a large customer base the company was losing millions of dollars each month. Using advanced analytics they were able to develop an attrition model to predict which customers were most likely to terminate their contract. In doing so, the company developed a model to cross-sell helping them to retain customers by providing products, services and other incentives targeted to their profile. This program improved the retention rate and contributed to an overall savings of $6.7 million.

You may not have yet heard the term "process mining" but it is a growing discipline with thriving new technology. Perhaps you have heard of data mining? Process mining is similar. Data mining is an analytical process using heuristics to explore large sets of data in search of consistent patterns and relationships. The goal of data mining is to be able to predict future behavior based on past activity.

Data Mining is an analytic process designed to explore data (usually large amounts of data - typically business or market related) in search of consistent patterns and/or systematic relationships between variables, and then to validate the findings by applying the detected patterns to new subsets of data. Data mining gained popularity as a business information management tool due to its predictive abilities, the results of which can help executives to make better business decisions.

OK, so what does that imply about process mining? Process mining enables the extraction of information from event logs. For example, the audit trails of a workflow management system or the transaction logs of an enterprise resource planning (ERP) system can be used to discover models describing processes, organizations, and products. The information in these logs represents a great wealth of untapped data. Event logs are ubiquitous in transactional information systems (e.g. WFM, ERP, CRM, SCM, and B2B systems) and until recently, the information in these event logs was rarely used to analyze the underlying processes.

So the basic premise behind process mining is to extract details from existing event logs to uncover patterns useful to the business. It is possible to uncover process, control, data, organizational, and social structures from event logs. And perhaps even more importantly, process mining can be used to monitor deviations from normal processing. Such activities are of paramount importance in the day-and-age of IT governance and regulatory compliance (such as required by the Sarbanes Oxley Act).

To be successful process mining requires so-called event logs and it's application is particularly useful in the context of workflow processes. A workflow process is the automation of a business process, in whole or part, during which documents, information or tasks are passed from one participant to another. The workflow is logged by the application that supports are automates the workflow. Then process mining techniques can be deployed to use the information collected in the log files to extract unexpected and useful knowledge about the process and then modify decision-making as appropriate in future instances.

There are many potential issues and problems that can be identified and corrected using processing mining. One such byproduct of process mining is simply identifying that critical activities are occurring as needed and when required. Additional solutions that can be arrived at using process mining include characterizing failures and successes. After finding the factors that lead to success steps can then be taken to optimize the workflow for future successes. When there are multiple choices within a workflow process mining can analyze the past choices to determine which more often led to desired results.

Process mining also can assist with general optimization. For example, process mining can help to identify redundant activities and operations which require restructuring. Moreover, process mining can be used to identify deviations from some desired process, e.g., some reference model or set of guidelines.

Example Applications

As we've established, the goal of process mining is to extract information about processes from transaction logs. Transaction logs hold a wealth of information across multiple types of applications, of which we will explore several examples. For process mining to be effective the information captured on the transaction logs must be of the following makeup:

1. each event refers to an activity (that is, a well-defined step in the process),
2. each event refers to a case (that is, a process instance),
3. each event can have a performer also referred to as originator (the person executing or initiating the activity), and
4. events have a timestamp and are totally ordered.

In addition events may have associated data (e.g., the outcome of a decision). Events are recorded in a so-called event log. A simple event log would look something like this:

Case ID	Activity ID	Originator	Timestamp
case 1	activity A	Joe	2005-03-27.15.01
case 2	activity A	Joe	2005-03-27.15.12
case 3	activity A	Elizabeth	2005-03-27.16.03
case 3	activity D	Carol	2005-03-27.16.07
case 1	activity B	Mike	2005-03-27.18.25

…and so on

This information can be used to extract knowledge. Many applications and systems produce transaction logs similar in nature to this.

Process mining is particularly useful in situations where events are recorded but there is no system enforcing people to work in a particular way. Consider for example a hospital where the diagnosis and treatment activities are recorded in the hospital information system, but where health-care professionals determine the "careflow."

A more common example is provided by an e-mail application such as Microsoft Outlook. An e-mail program is one of the most widely used software applications today. And such programs contain a rich source of information and processes to mine.

It is also possible to construct social networks from e-mail traffic. A social network is a set of social relations that connect people and groups; such networks can be examined for their impact on business operations and decision-making.

The challenge of process mining is to identify the case and the task for each event that is recorded. For example, given an e-mail message it is easy to see sender, receiver, timestamp, etc. However, if the e-mail is a step in some process, how to recognize the task and how to link the e-mail message to a specific case. Information such as threads, subject of the e-mail, and any special annotations can be used to extract meaningful event logs.

Enterprise resource planning applications such as SAP R/3 and Peoplesoft are also rich sources of process information that can be mined. However, the task of process mining such applications can be problematic.

SAP R/3, for example, creates many logs and reports. Unfortunately, the logs are either at a very detailed level or very specific for a given process. For example, reports such as the ST03 Transaction Report can be used to inspect database transactions. But these transactions are too fine-grained and do not point to a case and task. SAP R/3 also logs document flows which are more at the business level. As such, SAP R/3 can only be mined after considerable efforts because one needs to know the relevant tables and the structure of these tables to use the available document flows. This is not really a limitation of the concept of process mining but a result of the evolutionary growth of SAP R/3 resulting in a wide variety of logs requiring detailed business and technical knowledge to accurately utilize them.

ProM: A Framework for Process Mining

One example of a process mining implementation is the ProM (Process Mining) framework developed at Eindhoven University of Technology. The ProM framework provides a wide range of process miming techniques.

ProM has been developed as a platform for process mining algorithms and tools. Process mining aims at extracting information from event logs to capture the business process as it is being executed. Refer to Figure 1 for clarification. It provides an overview of process mining and the various relations between entities such as the information system, operational process, event logs and process models.

According to Wil van der Aalst, full professor at the Information Systems department of the Faculty of Technology Management of Eindhoven University of Technology, his team has used the ProM framework to mine several processes in practice, and have recently begun to mine hospital processes.

The original purpose for the ProM framework was to serve as a platform for process mining. As development ensued the scope of the framework grew broader to encompass tasks ranging from process verification to social network analysis to conformance checking and more. Additionally, the ProM framework supports a wide variety of process models enabling plug-ins to be added supporting additional models and operations.

For example, people can take transaction log from, say, IBM's WebSphere, transform it to MXML using ProM import, discover a process model in terms of a heuristics net, and convert the heuristics net to a Petri net for analysis. Such application scenarios are supported by ProM and demonstrate true model interoperability.

With respect to the applications discussed in the previous section, ProM can be deployed against Microsoft Outlook e-mail and SAP R/3 logs. In the context of the ProM framework it is possible to not only generate a social network from e-mail traffic but also process models. And the ProM framework has been deployed to apply process mining techniques to the various logs recorded by SAP R/3. (Application of the same approach to PeopleSoft is still under investigation.) But as mentioned earlier, current techniques have problems when mining processes that contain non-trivial constructs and/or when dealing with the presence of noise in the logs.

Professor van der Aalst notes that several techniques have been deployed to overcome these problems, including the use of genetic algorithms that are robust to noise. The professor goes on to say that "experiments show that the fitness measure leads to the mining of process models that can reproduce all the behavior in the log, but these mined models may also allow for extra behavior. In short, the current version of the genetic algorithm can already be used to mine process models, but future research is necessary to always ensure that the mined models do not allow for extra behavior."
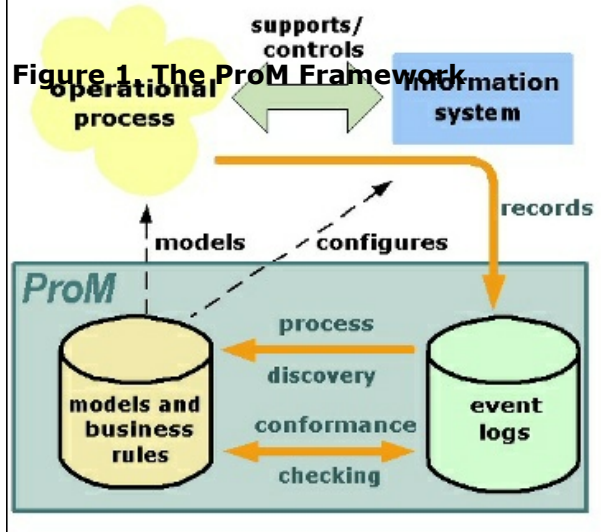
Bottom Line

The basic idea of process mining is to extract knowledge from event logs recorded by an information system. Using process mining techniques, the rich data resources just lying around in the transaction and workflow logs of popular application software can be turned into vital knowledge about your business operations. A thorough analysis of this information can greatly improve your business processes.

Figure 1. The ProM Framework

DB2PORTAL.com